# National Cancer Institute (NCI)
# Integrated Canine Data Commons (ICDC) Steering Committee (SC)

## Teleconference
## Wednesday, May 22, 2019

---

**Participants** (*Present)
External Committee Members
Matthew Breen*
Renee Chambers*
Dawn Duval*
Allison Heath*
Will Hendricks*
Warren Kibbe*
Debbie Knapp, ICDC-SC Chair*
Cheryl London*
Jeff Trent*
Roel Verhaak
Shaying Zhao*

Internal Committee Members (NCI, National Institutes of Health, and Frederick National Laboratory for Cancer Research [FNL])
Matthew Beyers*
Allen Dearry
Toby Hecht*
Paula Jacobs*
Tony Kerlavage*
Erika Kim*
Amy LeBlanc*
Christina Mazcko*
Phillip Musk*
Elaine Ostrander*
John Otridge*
Ralph Parchment, ICDC-SC Managing Secretary*
Connie Sommers*
Greg Tawa*

Others
Lori Lydard
Tara Whipp*
Mary Cerny (writer)*

**Opening and Welcome**

Dr. Parchment opened the meeting at 11:30 a.m. EDT and welcomed those in attendance.

**Minutes of the April Meeting**

The minutes of the April 10, 2019, ICDC-SC meeting were accepted as written.

**Data Governance Advisory Board (DGAB)**

The DGAB will oversee the development of the ICDC prototype. As not all data can or should be incorporated into the commons, the DGAB will advise NCI on which data and information to include, with the NCI Executive Team making final prioritization decisions.

Dr. Kibbe will serve as chair of the DGAB. Board members include:

- Four external (non-NIH) ICDC-SC members: Dr. Kibbe, Dr. Breen, Dr. Hendricks, and Dr. Verhaak
- Two internal (NIH) ICDC-SC members, including one member from the Center for Biomedical Informatics and Information Technology (CBIIT): Dr. Kim and Dr. Tawa

As reviewed during the April teleconference, a draft process for how data submission requests will be reviewed and submitted to the ICDC, and the role of the DGAB in that process, has been proposed.

**Best Practices Subcommittee (BPSC)**

Another team, the BPSC, has now been established for the ICDC. The BPSC will be responsible for identifying best practices for the ICDC to implement. The ICDC prototype will encompass clinical, pathological, genomic, biomarker, and imaging data as well as other components that will make up the data commons.

Dr. Trent will serve as chair of this subcommittee. Members include:

- External (non-NIH) ICDC-SC members: Dr. Trent, Dr. Chambers, Dr. Duval, Dr. Heath, Dr. London, Dr. Zhao
- Internal (NIH) ICDC-SC members: Dr. Jacobs, Dr. LeBlanc, Dr. Ostrander

**Top-Tier Questions, High-Priority Data, and Their Interrelatedness**

Several questions were posed to the Committee to begin the discussion:

**Canine Oncology**

- What are the landscapes of likely pathogenic germline and somatic mutations driving canine cancers? How do these landscapes correlate with other molecular features and clinical/demographic features of these cancers? How closely do these landscapes resemble those in human cancers of similar type? What about epigenomic, transcriptomic, proteomic features and tumor immuno/microenvironment, environmental history of canine cancers – comparing germline and somatic mutations
- Do germline and somatic cancer landscapes differ between breeds? Breed confirmation is important

- What is the evolutionary trajectory/natural history of canine cancer/individual dogs? How heterogeneous are these cancers? Interrelationship between drug effect and tumor evolution
- What mutational etiologies (mutation signatures) underlie canine cancer and what genetic lesions and/or environmental exposures are associated with these signatures?
- How to genomic signatures in the cancer change over time (natural progression, drug induced) and how do the changes affect drug resistance and cancer progression. What are the druggable targets?
- Using a bladder cancer focus, how do environmental risk factors and host factors affect the type of bladder cancer (subtype, immunotype) that develops? Response to therapy? Other malignancies?
- What factors (genetic, epigenetic, immune) can be identified that will predict the behavior of the cancer? Note that in humans and dogs ~50% of cases develop distant metastases and 50% do not. Currently it is not possible to predict the behavior in individuals. This means treatment is suboptimal.
- What genomic features are associated with sensitivity and resistance to therapy? i.e. genomic patterns that help determine which drug(s) should or should not be given to the individual. [this is going to be complex and will be tailored to specific therapy(s)]
- Can we build complex models that incorporate genomic, pharmacokinetic/dynamic, possibly proteomic and metabolomics data, treatment response, toxicities, etc. that can then be used in silico to predict outcomes with different therapies or with different host / tumor factors?

**Comparative Oncology**

- How closely do the landscapes of pathogenic germline and somatic mutation drivers resemble those in human cancers of similar type? What about similarities in epigenomic, transcriptomic, and proteomic features.
- What differences exist, if any, in oncogenic signaling pathways between dog and human? Do oncogenic mutations have different fitness effects in canine versus human?
- Using a bladder cancer focus, what are the molecular similarities (and differences) between invasive bladder cancer in dogs and humans (examples include molecular subtypes, specific gene and pathway variants, epigenetic changes, mutation signatures, tumor mutation burden, neoantigen load, immune cell responsiveness)?
- Why is superficial bladder cancer unusual in dogs when it is the most common form in humans?
- How to link the molecular findings (e.g., mutations, etc.) from a dog data set to those of its human counterpart?

1) What are the landscapes of likely pathogenic germline and somatic mutations driving canine cancers? How do these landscapes correlate with other molecular features and clinical/demographic features of these cancers? How closely do these landscapes resemble those in human cancers of similar type?

2) Same questions as #1, but for epigenomic, transcriptomic, and proteomic features of canine cancers.

3) What differences exist, if any, in oncogenic signaling pathways between dog and human? Do oncogenic mutations have different fitness effects in canine versus human?

4) Do somatic cancer landscapes differ between breeds?

5) What is the evolutionary trajectory of canine cancer? How heterogeneous are these cancers?

6) What mutational etiologies (mutation signatures) underlie canine cancer and what genetic lesions and/or environmental exposures are associated with these signatures?

A charge to the ICDC-SC is to identify the types of questions the commons will answer with the data imported into the ICDC as part of the larger NCI Cancer Research Data Commons (CRDC).

Two key areas under which these questions can be categorized include Canine Oncology and Comparative Oncology. In the first 3 years of the ICDC, as the prototype is developed and implemented, the Committee will be limited as to the types of test cases that can be used as questions. Although some questions will be answered using relatively small data sets, the Committee members were also asked to look ahead, "think big," and consider identifying questions as reflecting both shorter-term (e.g., 1-year) and longer-term (e.g., 10-year) goals of the commons.

The Committee discussed the following questions and issues under each category. For the purposes of this initial discussion, several questions use bladder cancer as a prototypical example. The questions will be expanded to include other malignancies, particularly those that are the most compelling for a specific issue or question. Suggested revisions to the questions are noted in italics, per the Committee's comments.

<u>Canine oncology</u>

- Bullet 1: What are the landscapes of likely pathogenic germline and somatic mutations that drive canine cancers? How do these correlate with other molecular features and clinical features? How do these landscapes resemble those in humans and epigenetic and proteomic features of canine cancers *as these data become available*?

  The Canine Comparative Oncology and Genomics Consortium (CCOGC) biospecimen repository includes transcription, genomic, proteomic, and copy number change data; pathology images; demographic and breed data; some outcome data; and proteomic and targeted resequencing of lesions. Data from this repository will be made available as the prototype is developed. To date, the repository includes five tumor types with 12 cases per type, or 60 cases in total. The five cancer types include osteosarcoma, melanoma, pulmonary carcinoma, and B- and T-cell lymphomas. For each type and case, there are clinical trial data, including pharmacokinetic/pharmacodynamic (PK/PD), tolerability, and response data. Clinical annotation for this data set is variable.

  More recently, activity data on drugs and synergistic drug combinations have been collected. The osteosarcoma cell line drug combination studies are finishing up this week, and initial

data will be available very soon. The drug combinations were identified based on molecular-level understanding gleaned from transcriptomic analysis of canine osteosarcoma samples. A larger trial of canine osteosarcomas has been approved and will be launched in the near future. Longer-range plans are to apply the approach used for the initial five cancers to the development of a clinically annotated specimen data set (for follow-up and standardized treatment) from dogs enrolled across the osteosarcoma trials, with as many as 100 cases for that set.

Data for the 60 cases were analyzed simultaneously. Results indicate five major gene code expression modules that underlie the data. The activation profiles of each of the modules are distinctly different. The highest level of commonality was found between the B- and T-cell lymphomas, which shared more of the underlying genetic underpinnings than the other cancers. The profiles are sufficiently distinct to be able to derive classification models to evaluate new data to determine cancer type. The data were run in triplicate (3 per each of the 12 cases), yielding good statistical power. It was noted, however, that even with 36 samples, the sample is still relatively small. The next step should be to validate the results on a larger sample. RNA and DNA were isolated from the same adjacent tissue. There are no remaining specimens, but DNA left over from the initial mutational profiling and the same tumor cell population used for expression analysis is available for whole-genome sequencing, pending funding.

The genomics of the cases for which transcriptome data are available include RNA-seq analysis of tissue samples to look at transcript expression. Analyses include mapping the transcripts to isolate the most relevant genes and then verifying that the protein products of those genes are also expressed in that tissue. The expression data identified about 80 genes. Follow-on proteomic studies show that the tissue proteomics profiles mirror the transcriptomic profiles in these dogs. Ultimately, these unique profiles may be useful in developing diagnostic tools. Testing to determine whether these profiles are present in blood is underway.

In looking for biomarkers for either diagnosis or targets for intervention, it is important to keep in mind that the work to date includes only five cancer types. Given this small sample, the specificity of any biomarker panel would be in question. The outcome could change as new data and additional cancers (and other disease states) are added to the analysis.

A question was posed as to the amount of variability in the response to the therapies given to treat these cancers in dogs, and whether correlations in genomic and transcriptomic profiles translate into consistency in response. It was noted that there are no therapeutic data for these dogs. The data and samples were collected as part of a multicenter banking effort. There were no standard or consistent treatments for these patients, only baseline and post-treatment assessments.

The Broad Institute has put together a curated list of genomes with somatic calls by breed. The list provides a very broad landscape of genomic variation across breeds (i.e., based on 80 million variants per the Institute's dog genome project). The list will be publicly available in the future.

Another data set on canine bladder cancer can be contributed to the ICDC. This data set includes clinical, PK, and pathology data.

Upon further discussion, the Committee considered whether the ICDC should be focused on a small number of cancer types or be opened to accept many different disease types from the start (i.e., with the prototype) to begin to compare the genomic landscape across diseases. Committee members generally favored the latter approach, especially with regard to the informatics tools being developed for the canine commons. These tools are most powerful when they are able to compare and contrast disease states to arrive at gene modules. Further, looking more broadly outside cancer can help create more specific descriptions of cancer types by clarifying what a cancer is not. Simultaneous analysis of cancer and other diseases starting in the prototype would build a strong foundation for the ICDC.

Another question involves the different technologies that measure different aspects of tumor biology. The problem is identifying or developing the tools needed to knit together disparate pieces of data into a common home. For example, what tools are needed to seamlessly relate comparative genome hybridization data to RNA-seq data to mutation data? This is a broader and more ubiquitous problem with a lot of overlap across questions, and the problem is not limited to canine oncology research. Methodologies will be validated as more data are collected and integrated. The underlying questions for this issue might be identifying methodological gaps for this field of research and determining what data sets are needed to develop the methodologies to close those gaps.

In looking at the microenvironment and other factors, a series of best practice tools will be part of the Committee's ongoing discussions. For example, single-cell sequencing is used to establish cell transcriptome in the microenvironment and address heterogeneity questions.

The Committee decided that it would identify relevant questions and that it is premature to prioritize questions at this point. The questions will direct the type of data needed. Prioritization of the questions can be done at a later time, as the amount and type of data available vs. still needed become clearer.

The Committee found the set of questions listed under the first bullet to be relevant.

- Bullet 2: The question is relevant but should be revised to include *with comparing germline and somatic mutations* and *tumor microenvironment and environmental history of cancer in individual dogs and different breeds of dogs*. A reference to *confirmation of breed* should also be added.

- Bullet 3: In cases and studies in which sequential biopsies are done over time (e.g., pretreatment and post-treatment), there could be different mutations and different drivers of the disease that, in turn, affect the evolution of the tumor. The question for this bullet is general and needs to be revised to take these factors into consideration. Specific issues include understanding the natural history and trajectory of a particular cancer in an individual dog and how the cancer trajectory is affected by different treatments. This bullet also

addresses questions regarding heterogeneity of cancer in humans, which are the same for dogs (e.g., selecting for existing clones in looking at resistance, promotion of rapid evolution of resistant clones and tissue, how metastases differ from each other and the primary tumor). The Committee suggested revising this bullet to reflect interest in how drugs alter the cancer trajectory as well as how the evolution of the tumor affects the drug response. The added language included *interrelationships between drug effects and tumor evolution in different dogs and different breeds*…

- Bullet 6: Committee members recognized that the data sets and questions posed go beyond bladder cancer and include other cancers. The wording *and other malignancies* should be added to the end of the last statement under this bullet.

- Bullet 7: Although these points also relate to bladder cancer in humans, it is not clear how to predict how certain cancers behave. In both humans and dogs, approximately half of bladder cancers progress to metastatic disease, while half respond to treatment. Disease progression in bladder cancer appears to be related to molecular subtypes, but the predictive value of this characteristic has not been confirmed and warrants further investigation.

- The last bullet/question in this category relates to methodological gaps, including how to build models to predict outcomes even before the appropriate data are available. This is an overarching issue, and addressing this question is more likely a longer-term goal. Committee members suggested adding *and understand mechanisms, including disease resistance mechanisms* after "… building complex models to predict outcomes…"

There is considerable overlap in the questions posed for the canine oncology category. It was noted that many U01s include clinical trials in dogs. Researchers with specimens from responders should be encouraged to submit these samples to the ICDC; those samples can then be used to determine why certain treatments are effective in certain animals with the goal of developing biomarkers of response.

<u>Comparative oncology</u>
Several of the questions and issues in this category are similar to those in the canine oncology category. The Committee reiterated the point of thinking beyond the specific example of bladder cancer, which is used to address several questions posed regarding comparative oncology. For example, it was noted that gliomas, like bladder cancers, are classified according to molecular subtypes.

The Committee discussed whether data need to be collected and organized in a way to allow for rapid analysis in comparison with publicly available human data sets, unless this is already the plan for how the commons will be set up.

A key operational aspect of the commons is to have the appropriate terminologies and ontologies in place for queries that give the closest alignment between human and canine data. It was noted, for example, that descriptions of diagnoses in people are not consistent and would need to be addressed. If the same problem exists in veterinary medicine, then further clarification or standardization would similarly be needed. Gene identification can be straightforward in some

diagnoses but does not necessarily address the issue of variants. The National Center for Advancing Translational Sciences and projects such as the Monarch Initiative are in part aligned to perform comparative oncology and could serve as resources for this component of the ICDC.

Using bladder cancer as an example, it was noted that the majority of invasive bladder cancer is the same histological type, but due to changes in terminology over the past few years, the vocabulary for classification based on histology is not consistent. Having histological images and/or routine pathology reviews to confirm mainline histological types may be the means to standardized diagnoses to assure consistency regardless of changes in terminology. The Cancer Imaging Archive could serve as a model for this aspect of the commons.

Canine vs. comparative oncology

Committee members discussed whether the ICDC should focus on canine oncology first to populate the prototype and then build on that foundation with comparisons to human cancers, which will require interoperability between the ICDC and the CRDC; or whether canine and comparative oncology components should be developed in parallel in building the prototype.

The Committee revisited the primary and long-term goal of the ICDC, which is to determine whether pet dogs with spontaneous cancers can serve as close models of human disease in order to evaluate new drugs, immunotherapeutic agents, and combination therapies for further development for human cancer patients. The research questions for the ICDC should be delineated within this context so that those who are contributing to the commons have a clear understanding of these questions to assure that they collect the appropriate data and specimens. The ICDC is a very valuable resource, and it is important to build the comparative oncology component from the start to make sure collection of important data and samples is not neglected or forgotten until much later in the process. Thus the Committee should be cognizant of the comparative value of the commons at all points in the prototype phase and beyond, to take advantage of the knowledge on both sides of the project.

Another point to consider is the unique opportunities canine models can provide that will add value to the ICDC and what questions those opportunities can answer. For example, dogs are immunocompetent and in that respect are closely related to humans. The ICDC could be used to develop immunocompetent models for development and testing of therapies for canine and human patients. Dogs also live in the same environment as people and are exposed to the same things (e.g., secondhand smoke, air pollution).

A relevant question is what further information or key pieces of data are needed to help better understand the immune environment in these cases (e.g., CD3 immunohistochemistry on tumors) and follow the response over time. Investigators, including ICDC-SC members who have U01s, which include immunotherapy studies, can serve as resources and guides for these questions. The SC may want to consider establishing a subcommittee in the future that will focus on immunotherapeutic issues.

Breed-specific information

Caution is needed when identifying and confirming breed. It was pointed out that breeds listed on specimen labels are not always correct, particularly if the breed is “determined” by the dog’s

appearance, a relatively common practice. SNP analyses of breed-specific markers of "outliers" indicate that dogs thought to be a certain breed are often mixed-breed and not purebred dogs. To assure the accuracy of the breed of the animal being studied and therefore optimize data analyses, the ICDC should have a confirmatory method or metric in place. The genomes of hundreds of breeds of dogs have now been sequenced and validated. In addition, there are numerous breed registries for which parentage or lineage has been verified to confirm breed. For example, American Kennel Club breed confirmation is about 80–90% accurate. Genetic profiles with breed-specific SNPs are also available to confirm breed or breed mixes. A 670K SNP chip has been developed based on analysis of hundreds of whole-genome sequences that identified about 20 million SNPs, which were then narrowed down to the most informative SNPs for confirming breed. It was noted that in the United States the most common breed that develops cancer is the affenpinscher. The ICDC could specify that breed be assigned based on genomics, DNA, and blood. A more general guideline would be to encourage veterinarians to label the breed as "unknown" unless the breed is genomically confirmed.

Other questions/types of data to collect

The dog is recognized as a valuable model for this research enterprise and will complement but not replace other animal models. However, dogs can be especially helpful in elucidating markers and testing therapies that cannot readily be done in humans. In the bladder cancer field, for example, it is difficult to test drugs in humans that do not already have a relatively well-established safety profile unless there are compelling data in a relevant model. The ICDC could fill this niche, whether for bladder or other cancers.

One of the fundamental challenges in human cancer treatment with respect to testing of novel/investigational combination therapies is that they are usually relegated to trials of patients with advanced or terminal disease. Canine models might provide a way to explore novel test therapies at earlier stages, such as with microscopic disease, where there is an opportunity for a better response. Having genomic data will add even more value to such studies by advancing understanding of how some of these therapies work, particularly in treatment-naïve patients vs. those who are refractory to standard treatments. It was noted that there are very few standards of care for cancer treatment for dogs. As a result, it is not difficult to get access to investigational drugs at earlier stages of the disease, when a dog's immune system is more robust.

Another area where the ICDC could have an impact is as an animal counterpart for rapid autopsy programs for people. In humans, such programs have been slow to start, and most autopsies are performed at least 2 hours after death. In contrast, in veterinary medicine programs, with the consent of pet owners, autopsies are usually performed within 15 minutes of death.

Additional information was requested about software tools that can be used in canine cancer research, and how or whether such tools used in humans can be translated for use in dogs. It was noted that a range of tools exist, some of which can be directly applied to dogs and some of which need to be adapted and optimized for use with dogs. In addition, there is a suite of tools that can be used for best practices. Such tools can be part of an ongoing discussion on defined analysis for the BPSC and the DGAB. Other teams are also working on synergizing the animal/veterinary data pipeline to create a common platform for analysis.

Another issue for consideration is whether research questions in humans can address issues specific to cancer in dogs.

The Committee will continue to continue to explore what questions can be asked about cancer in dogs that cannot be investigated in people to maximize the impact of the ICDC on human disease.

Short- vs. long-term goals

Addressing the question of short- vs. long-term goals requires an understanding of how to obtain the correct data set(s) for the commons, including the quantity of data needed, the number of animals needed to answer the questions being asked, and whether there is a human tumor/cancer for comparison. Other factors include the design of the data sets and the experiments run to obtain those data. Although some data sets are currently available or soon will be, a challenge is that it is not possible to know what data sets will become available. This further complicates the question of whether an issue should be designated as part of a shorter- or longer-term goal.

**Administrative Items**

June meeting

The next meeting of the ICDC-SC will convene via teleconference on Wednesday, June 26, 2019, from 11:30 a.m. to 1 p.m. EDT.

Intellectual property policy

Per internal discussion within NCI, it is unlikely that creation of intellectual property will be part of the work of the ICDC-SC, given that the goal of the commons is to create a product that is publicly available. To that end, a statement as to the policy for intellectual property for the ICDC-SC has been drafted, as follows:

## Intellectual Property Policy

The work of the SC is advisory to the NCI, so it should not need to create any IP to perform its function. Its role is to advise the NCI on how best to proceed with the build and the data content of a ICDC prototype. In addition, an important feature of the data commons is that it will be open to the research community so that data can be democratized, meaning everyone in the research community can access and analyze the data without restrictions on its use, user fees or other limitations. It would be consistent with both the mission of the NCI and the purpose of the program to designate any analytical tools developed for use in conjunction with the research commons as “research tools” for which the NCI would not pursue patents and would make freely available to the research community.

In addition, there was brief discussion reiterating the important of the NDA/COI agreements that the ICDC members have signed:

> **Non-Disclosure Agreement and Treatment of Data Sets**
>
> Also, the confidential treatment of submitted data sets by the SC when considering them for inclusion in the prototype canine data commons is very important for earning the trust of investigators who will submit their data sets for consideration. The NDA specifies that the only permitted use of submitted data is to conduct the work of the SC, and that these submitted data sets cannot be distributed to anyone else.

Conflict of interest (COI) and honoraria

All external Committee members have now submitted their COI disclosure forms for participation in the ICDC-SC. Dr. Parchment thanked the Committee members for providing this information.

**Action Items**

- Dr. Parchment will distribute the list of questions reviewed during the meeting.
- Mr. Beyers will follow up with the chairs of the DGAB and BPSC to coordinate future meetings/teleconferences (TBD).
- Committee members were asked to forward additional questions for the ICDC to Dr. Knapp, who will distribute the questions to Committee members for comment in between teleconferences.
- Topics for future meetings should be forwarded to Dr. Knapp, Dr. Parchment, Dr. Hecht, or Mr. Beyers.

**Adjournment**

The meeting was adjourned at 12:50 p.m. EDT.